7.0 GENERIC ISSUES ON EPIDEMIOLOGICAL EVIDENCE

In the DHS Risk Evaluation Guidelines (see Appendix 2) the three reviewers proposed to organize their pro and con arguments around a series of pre-specified questions relevant to developing a degree of confidence as to whether epidemiological associations were causal in nature. Because these factual issues are also relevant to policy, they developed questions relevant to the status of research assessing dose-response relationships, any unequal vulnerability to EMFs, or an unequal distribution of exposure. The questions in the Guidelines are summarized by the questions in the following two tables, and these are repeated for each endpoint specifically considered. Having pre-specified questions such as these assures a systematic evaluation.

Following the scheme of IARC, the reviewers first asked (see Table 7.1) if the associations observed could be due to chance, bias, or confounding. If not, they systematically examined attributes of the evidence which might incline us to attribute the association to causation.

As the reviewers went through the specific diseases using these standard questions, they realized that some of them always involved the same pro and con arguments and that they always came down on one side of the argument, regardless of the disease being considered. They decided to deal with those questions in this section and only mention them in the summary tables for the respective diseases.

TABLE 7.1 QUESTIONS RELEVANT TO CAUSALITY

EXPLANATIONS OF A STATISTICAL ASSOCIATION OTHER THAN A CAUSAL ONE
Chance: How likely is it that the combined association from all the studies of EMF and disease is due to chance alone?
*Bias: How convinced are the reviewers that EMFs rather than a study flaw that can be **specified and demonstrated** caused this evidentiary pattern? If no specified and demonstrated bias explains it, how convinced are they that EMFs caused these associations rather than **unspecified** flaws?*
*Confounding: How convinced are the reviewers that these disease associations are due to EMFs rather than to another **specified and demonstrated** risk factor associated with EMF exposure? If not due to a specified risk factor, how convinced are they that they are due to EMFs rather than to **unspecified** risk factors?*
*Combined effect: How convinced are the reviewers that these disease associations are due to EMFs rather than to a combined effect of chance and specified or **unspecified** sources of bias and confounders?*
ATTRIBUTES SIMILAR TO HILL'S (Hill, 1965) THAT ARE SOMETIMES USED BY EPIDEMIOLOGISTS TO EVALUATE THE CREDIBILITY OF A HYPOTHESIS WHEN NO DIRECT EVIDENCE OF CONFOUNDING OR BIAS EXISTS
Strength of association: How likely is it that the meta-analytic association is strong enough to be causal rather than due to unspecified minor study flaws or confounders?
Consistency: Do most of the studies suggest some added risk from EMFs? How likely is it that the proportion of studies with risk ratios above or below 1.0 arose from chance alone?
Homogeneity: If a large proportion of the studies have risk ratios that are either above or below 1.0, is their magnitude similar (homogeneous) or is the size of the observed effect quite variable (heterogeneous)?
Dose response: How clear is it that disease risk increases steadily with dose? What would be expected under causality? Under chance, bias, or confounding?
Coherence/visibility: How coherent is the story told by the pattern of associations within studies? If a surrogate measure shows an association, does a better measurement strengthen that association? Is the association stronger in groups where it is predicted? What would be expected under causality? Under chance, bias, or confounding? How

convinced are the reviewers that the magnitude of epidemiological results is consistent with temporal or geographic trends?
Experimental evidence: How convincing are the experimental pathology studies supporting the epidemiological evidence? What would be expected under causality, bias, chance, or confounding?
Plausibility: How convincing is the mechanistic research on plausible biological mechanisms leading from exposure to this disease? What would be expected under causality, chance, bias, or confounding? How influential are other experimental studies (both in vivo and in vitro) that speak to the ability of EMFs to produce effects at low dose?
Analogy: How good an analogy can the reviewers find with similar agents that have been shown to lead to similar diseases? What would be expected under causality, chance, bias, or confounding?
Temporality: How convinced are the reviewers that EMF exposure precedes onset of disease and that disease status did not lead to a change in exposure?
Specificity and other disease associations: How predominantly are EMFs associated with one disease or subtypes of several diseases? What would the reviewers expect under causality, chance, bias, or confounding? How much is their confidence in EMF causality for disease X influenced by their confidence that EMFs cause disease Y?

The reviewers next asked (see Table 7.2) questions relevant to dose response and policy, including factual questions relevant to the environmental justice policy perspective and questions about the current state of science in the area. In many cases, however, the evidence is insufficient to provide an answer.

TABLE 7.2 QUESTIONS RELEVANT TO DOSE RESPONSE AND POLICY

How confident are the reviewers that a specific exposure metric or aspect, other than 60 Hz TWA magnetic field, is associated with this disease?
How confident are the reviewers of evidence for threshold or plateau?
How confident are the reviewers of evidence for biological windows of vulnerability?
How confident are the reviewers of a consistent induction period or required duration of exposure?
How does EMF compare to other risk factors for this disease, as to added risk to the total population and to highly exposed people?
How does the observed relative risk compare to that which would generate a 1/1000 or 1/100,000 theoretical lifetime risk?
How confident are the reviewers of evidence for racial, gender, or class differences in exposure or vulnerability? (This is relevant to environmental justice.)

State-of-science questions.

How much room for improvement in quality or size is there in the best existing studies?
How many new studies are in the pipeline and how capable are they of changing the reviewers assessments?
How likely is it that further studies could resolve controversies?

7.2 Approaches to Weighing Streams of Evidence

The reader will notice that, following Hutchison and Lane (Hutchinson, 1980), the three reviewers have phrased these questions so that they would be answered in a graded fashion rather than in a "yes" or "no." They have been worded so that when the reviewers answer with a larger likelihood or degree of confidence, this means that the strength of evidence for causality has increased. This is helpful in thinking about the weight to be given to the answer and in avoiding the pitfall of simply adding "yes" and "no" answers. Following Hutchison and Lane's recommendation of "etiological balancing," many of these questions can be conceptualized by comparing the likelihood of the pattern of evidence (if EMFs really caused the disease in question) to the likelihood of the same evidentiary pattern, if only chance, bias, or confounding had produced the pattern of evidence. So, when the reviewers ask themselves about bias, they couch it as their convictions about EMF causality relative to their convictions about the presence of specified or unspecified study biases. An exception is the question about chance, where the conventional question is posed about the likelihood of the pattern of evidence under the null hypothesis.

In DHS's Risk Evaluation Guidelines, the reviewers pointed out that the *size* of the relative likelihood conveyed by supportive or unsupportive patterns of evidence depended on 1) how good that stream of evidence was in detecting a cause, if it usually detected a harmful agent (sensitivity); and 2) how good that stream of evidence was in not falsely implicating an agent (specificity). The reviewers pointed out that unsupportive patterns of evidence from a stream of evidence that often missed detecting a cause did not pull their confidence down very much, and that supportive patterns of evidence from a stream of evidence that often falsely implicated agents would not pull confidence up much. (See pages 48–52 of Appendix 2.)

As a heuristic, the reviewers can think of the size of these relative likelihoods as the weights given to the different streams of evidence. For example, the question, "How clear is it that risk increases steadily with dose?" could be rephrased as, "How much more or less likely is the observed dose response pattern if EMFs caused disease X than if chance, bias, or confounding had produced this pattern?" Suppose that, in studies where few subjects have high exposures, an inconsistent dose-response pattern might be expected under the EMF hypothesis, and that this is somewhat more likely to be seen than if only chance, bias, and confounding were at work. This pattern of evidence would then increase confidence somewhat, and the heuristic relative likelihood would be a number bigger than one.

Of course, the answers to these questions cannot be mechanically considered in isolation. Certain combinations of answers influence the reviewers degree of confidence more than the isolated answers would predict. For example, one might be quite sure of a minor bias at work in all of the studies, but if the those studies all reported relative risks of 20 with tight confidence limits, concerns about bias would not weigh as highly as would be the case if the studies all reported relative risks of 1.1. That is why the reviewers had to consider the pro and con answers to the structured questions and then come to an integrated judgment about what the evidence suggested, rather than assigning scores and mechanically multiplying them or adding them up.

7.3 General Points About the Causality – Relevant Questions

The reviewers found that some of the questions were harder to formulate in the relative likelihood mode. So, in this section, they have explained how they approached those questions.

Chance

The question about chance simply asks how probable the observed, or a more extreme, pattern of evidence is under the null hypothesis of "no association." If it is quite probable (say 6 times out of 100) under the null hypothesis, then conventional thinking dismisses the pattern of evidence as being due to chance. The DHS reviewers ask this question of the pattern of relative risks and of meta-analytic estimates of effect because IARC specifically considers this. Since it is conventional to do so, decision makers may choose to pay attention to how likely the evidence is under the chance hypothesis. A pattern unlikely under the null hypothesis could be interpreted as follows: "If these were randomized experiments without the possibility of bias or confounding, the statistical associations found would not be expected to occur by chance in 5 or fewer experiments out of 100 replications, if there was really no effect." Of course, epidemiological studies are not experiments. It would be unethical and impractical to experimentally subject large numbers of humans to potentially harmful agents. This leads to the consideration of bias and confounding.

Bias

Any source of error in collecting the data may introduce a bias, which is a reason why the apparent result might not be the truth. A very common bias results from errors in assessing the true exposure of the subjects to the agent of interest, in this case EMFs. Provided exposure of cancer cases and healthy controls is not

assessed differently, this bias on average results in an underestimate of the risk, if one exists. When comparing the health risk of subjects exposed above one value to that of subjects below that value, non-differential misclassification of exposure* would not, on average, show an association if one does not truly exist. However, it may inflate the risk of intermediate exposure subjects and thus frustrate attempts to estimate a dose-response function. In most of the EMF studies, measurements were not taken for a long enough duration during the induction period of the disease to avoid this kind of misclassification. And there is even some argument about whether the right aspect of the EMF mixture has been measured. The three reviewers concluded that all of this may have led to an underestimate of any true effect of high versus low exposures and may have frustrated the ability to develop an appropriate dose-response curve.

Of the many errors that can creep into epidemiological studies, one in particular has been a source of argument with regard to a subset of the EMF epidemiological studies. The reviewers refer to “selection bias” in some of the case control studies. A case control study is analyzed by comparing a series of cases with a disease to a series of healthy subjects as to their EMF exposure. If the cases display a higher proportion of high EMF exposure than the controls, this suggests a causal effect of EMFs. If, however, the probability of being selected for study is influenced both by whether one has the disease AND whether one had a high EMF exposure, then an apparent difference will appear between the cases and the healthy controls, which is the result of this biased selection and the result does not reflect any true effect of EMFs on the disease. One way to recruit healthy subjects is random telephone contact. This method excludes subjects of lower socio-economic status (SES), who may not have a telephone. Experience has shown that healthy controls of lower SES are sometimes less likely to participate in epidemiological studies than upper class subjects. In some studies, lower class subjects are more likely to live in neighborhoods with nearby power lines (Bracken et al., 1998). Since cancer patients of all social classes are easier to recruit (through a cancer registry) and more likely to be interested in participating, the effects of non-representative control selection may distort the comparisons between cases and controls and, therefore, the study results. In the case of EMF, it is claimed that the fact that there are more subjects living close to power lines among the cancer patients than among the healthy controls could be due to the fact that low SES subjects are more likely to live close to power lines and they are underrepresented in the control group. This issue of possible selection bias in case control studies is a particular issue for the North American case control studies on childhood leukemia. Hatch (Hatch et al., 2000) indicate that the association between childhood acute lymphoblastic leukemia (ALL) and front door magnetic fields greater than 3 mG was 1.9 (1.1-3.27) among full participants in their study but fell to 1.6 (0.98-2.61) when 147 partial participants were included. Although this difference was well within sampling variability, she suggested that it might be evidence of the presence of a selection bias which might be even more extreme if non-participants had their front doors measured and had been included in the analysis. Hatch (Hatch et al., 2000) concluded that "while confounding alone is unlikely to be an important source of bias....selection bias may be more of a concern...in case-control studies." The Scandinavian studies relied on cancer registries and lists of citizens and did not require permission of the subjects so that selection bias was not a problem. Ahlbom (2001) has shown that the results of the two groups of studies are not much different. The pooled analysis of all the studies he dealt with showed a relative risk for exposures above 4 mG as 2.0 (1.3-3.1), while the results after excluding the US studies was 1.7 (1.0-2.8). That is, the confidence interval of the two risk estimates overlap, indicating that there may or may not be some over-estimate of the effect of living near power lines in the American studies, but that even if these are excluded, the association remains statistically significant. In the pooled analysis by Greenland et al. (2001), there was an effect of power line proximity (“wire code”), as well as an effect of measured magnetic fields. This might indicate some selection bias for power line proximity. Nonetheless, magnetic fields come only partially from power lines. Internal wiring and currents on plumbing form an important source (Zaffanella & Kalton, 1998). The only evidence we know of that examines personal EMF exposure from all sources and its relation to social class (Lee GM & Li D-K, personal communication) does not suggest differences in personal EMF exposure in different social classes. The evidence linking EMFs and adult leukemia, adult brain cancer, Lou Gehrig’s Disease, and Li’s prospective miscarriage study come largely from study designs where selection bias is not possible (studies where rosters of healthy workers or subjects of high and low exposure are followed until death or health outcomes are determined from available records without requiring subject cooperation). Thus, although selection bias may have distorted the associations between EMF and childhood leukemia in some of the studies, the three reviewers did not believe that it totally explained the childhood leukemia findings and selection bias was not even an issue in the bulk of the studies related to adult leukemia, adult brain cancer, ALS, or in one of the two recent studies on EMF and miscarriage.

CONFOUNDING

The term “confounding” is derived from the Latin “confundere,” to melt together. Epidemiologists use the term when the impact of two risk factors “melt together” and

* "non-differential misclassification of exposure" is said to occur when errors of measurement occur equally in cases of disease and in healthy controls.

must be disentangled. If heavy alcohol consumption and smoking are both known to cause esophageal cancer, and people who drink also tend to smoke, then the effect of drinking will confound the effect of smoking and vice versa. Therefore, one must correct for this confounding in the way the data are analyzed. Sometimes the non-effect of a factor which conveys no risk at all is confounded with the true effect of another factor. For example, it has been suggested that people who live near power lines also live on busy streets with lots of traffic and air pollution. This argument suggests that the effect of air pollution on childhood leukemia was confounded with the non-effect of the power lines, and the power lines were falsely implicated instead of the air pollution. Two conditions must pertain for an agent to be a strong confounder of the EMF effect on the various diseases discussed in this report. That agent must be strongly correlated with EMF exposure and it must have an effect on the studied disease that is even stronger than the apparent effect of EMF. If it is weakly correlated with EMF exposure it must have an effect on disease that is very strong indeed if it is to make EMF falsely appear to have an effect. Langholz (Langholz, 2001) has examined the candidate confounders for childhood leukemia and their association with wire code. He concluded that while something connected with the age of home was a possibility, factors like traffic density, ethnicity, and smoking were not likely confounders. Indeed, not all studies of traffic and childhood leukemia suggest it as a risk factor (Reynolds et al., 2001), but a recent study of traffic and power line proximity and childhood leukemia (Pearson et al., 2000) did suggest that there might be a joint effect. Hatch (Hatch et al., 2000) examined a variety of socioeconomic, and other confounders, and concluded that together, or alone, measured confounders would distort the association with ALL by less than 15%. Hatch also found no association between residential mobility, magnetic fields, or leukemia unlike Jones (Jones et al., 1993).

Electric shocks have been invoked to explain the relation between high-exposure jobs in the utility industry and ALS (Ahlbom, 2001), (NRPB, 2001a). If this were confirmed, they might also be invoked to explain the adult leukemia and brain cancer associations on the as yet unproven assumption that shocks could somehow cause cancer. However, the literature linking shock to ALS, unlike much of the literature linking high-EMF exposure jobs to ALS, depends on subjects remembering shocks. They are thus more vulnerable to recall bias than the EMF studies. Some of the studies suggest a protective, not a harmful, effect (Cruz et al., 1999), (Kondo & Tsubaki, 1981), (Gunnarson et al., 1992) and the size of the harmful effects of shock are less than the high EMF job effect (Deapen & Henderson, 1986), (Savettieri et al., 1991). No published study has demonstrated a correlation between shocks and high-EMF exposure jobs. Studies are underway to see if grounding currents are associated with measured magnetic fields and power line proximity. The three reviewers felt that the evidence for the confounders that had been proposed for EMF exposure did not have strong support and therefore their degree of confidence was not decreased by the pattern of evidence.

COMBINED EFFECT OF CHANCE, BIAS, AND CONFOUNDING

Although each of these possibilities by itself is unlikely to explain the association between EMF and cancer, is it possible that a combination of the three may be responsible for an artefactual finding? The DHS reviewers considered this possibility and concluded that this is not a credible explanation when many studies of different design have reported similar results. It is not impossible that individual studies may be have their result completely explained by an extraordinary coincidence in which independent unlikely events occur simultaneously. However, for many diseases considered here the general pattern of results is not critically dependent on accepting each individual study as reliable. For example, in the case of childhood leukemia, it has been repeatedly shown that, even if a few studies are excluded, the results of meta-analyses, pooled analyses, or sign tests are not significantly altered.

In conclusion, the DHS reviewers, to different degrees, concluded that chance, bias, and confounding are not probable explanations for the reported associations when they have been reported repeatedly by independent investigators. In addition, the DHS reviewers considered other criteria, notably Hill's criteria for causality, keeping in mind that these are not to be considered as strict rules to follow. Apart from consistency, which, as noted above made them doubt the non-causal explanation for a few endpoints, none of the Hill's attributes, when applied to the pattern of evidence, influenced their degree of certainty by much.

The DHS reviewers recognize the size of the associations between EMF exposure and the various diseases studied are not so far above the resolution power of the studies that confounding and bias could be definitively ruled out as explanations. They recognized that there was rarely an orderly progression of increased risk within studies and that the effects reported for groups with dramatically high exposures like electric train operators did not display dramatically high risks when compared to those with low or moderate exposures. There are also examples where the statistical results are not completely coherent. However, these evidentiary tests are prone to giving false-negative results due to non-differential measurement error and sample size problems. Also, EMFs may have societally important effects that are nonetheless truly close to the detection of epidemiology. Finally, an agent may act in an "on/off" fashion and would not produce a steadily increased effect. These patterns of evidence therefore lowered confidence some, but not a lot.

STRENGTH OF ASSOCIATION

As the apparent relative risk conveyed by EMF exposure gets further and further away from 1.00, the likelihood of the pattern occurring under chance gets smaller and smaller. Prior experience with research studies suggests that, if specific evidence for particular bias or confounding is not present, the probability of unidentified bias or confounding falsely producing an apparently harmful or beneficial association gets smaller and smaller as the association moves away from the null value of RR = 1.0. This means that the likelihood of the evidence under causality RELATIVE to the likelihood of the evidence under bias, confounding, or chance gets bigger and bigger as the relative risk departs from 1.0. However, the posterior probability does not necessarily become greater as the relative risk increases. For example, all three core reviewers had a vanishingly small prior probability that residential EMFs could increase the risk of various diseases 100-fold because this would already have been noticed. If there were an epidemiologically detectable effect, they thought it would be found in the range of relative risks between 1.2 and 5. So, if the reviewers observed a relative risk of 100 in a particular study, their posterior would be less than if they observed a relative risk of 2.00. Some of the core reviewers took the position that a small RR simply did not support the causal hypothesis very strongly but did not go against the causal hypothesis. Other core reviewers gave somewhat more weight to the bias considerations if the pooled RR for the various studies was close to 1.0.

CONSISTENCY

"Consistency" refers to the consistency of the results with the hypothesis of an EMF risk (the reviewers refer to the consistency between studies as "homogeneity"—see below). This concept is useful if the body of evidence consists of a fair number of studies. The reviewers ask if the proportion of studies with risk ratios falling above a relative risk of 1.0 could easily be due to chance, by calculating the cumulative binomial probability of the observed number of risk ratios above a RR of 1.0. If they are nearly equally distributed above and below a RR of 1.0, then the results are not consistent. If all or most are above or are below a RR of 1.0, then the results are consistent. Consistency is hard to evaluate when there are only a few studies. Suppose the body of evidence contained only one large and one small study, each showing a RR above 1.0, and one small study showing a RR slightly below 1.0. The meta-analysis in this case might suggest a statistically significant association above a RR of 1.0. In that case, the pattern of the three risk ratios might easily seem to be randomly inconsistent because of the small number of studies, even though 66% of the studies were above a RR of 1.0. The reviewers recognize that for endpoints in which all the studies had been subjected to a meta-analysis or pooled analysis, a more elegant way to assess what is referred to as "consistency" and "homogeneity" would be to analyze the components of variance around the summary measure of association. This kind of information was not usually available to the reviewers and they attended to the proportion of relative risks above and below unity, as an approximate way of characterizing the evidence.

HOMOGENEITY

Even if the relative risks in a series of studies were consistently above a RR of 1.0, their sizes might not be homogeneous. For example, women with a particular gene might have a large risk of a birth defect from smoking while women without that gene might have a much smaller effect. This would produce a pattern of relative risks between the smoking habit and the birth defect that was consistent but not homogeneous.

EXPERIMENTAL EVIDENCE (ANIMAL PATHOLOGY)

The reviewers agreed that, with few exceptions, animal pathology studies based on high exposures to certain aspects of the EMF mixture showed no effects. There were three reasons why the reviewers believed that animal bioassays of single ingredients of the EMF mixture might be prone to missing a true effect:

a) Finding the right animal species to test: While the reviewers recognized that most agents found to cause cancer in humans also cause cancer in some (but not all) animal species, they were also cognizant that there are known human carcinogens, such as cigarette smoke, alcoholic beverages, benzene, and arsenic for which no animal model existed for many decades.

b) Testing one ingredient of a mixture: The reviewers all questioned whether the bioassay of one element of a mixture could be sensitive enough to detect problems in the entire mixture. For example, many reassuring assays on the carcinogenicity of caffeine would not reassure them about the carcinogenicity of coffee. The animal pathology studies to date have been on pure steady 60 Hz fields not on the mixture of ingredients found near power lines or appliances.

c) Assuming that high intensities of magnetic fields produce larger effects than moderate fields do: The reviewers also questioned the sensitivity of a bioassay involving a small number of animals and assuming a monotonically increasing risk from low to high dose, when the epidemiological studies that prompted the bioassays did not suggest an ever-increasing response.

The epidemiology suggests there is either no effect at all (Tynes, Jynge & Vistnes, 1994a) or no more effect at 250 mG (Minder & Pfluger, 2001) than 3 mG in children (Greenland et al., 2000), or 24 hr TWA of 7 mG in highly exposed utility workers (Kheifets et al., 1997b), (Kheifets, 2001). One would not expect rodents at 1000 mG to demonstrate a large enough effect to be detected in a conventionally sized laboratory experiment with a few hundred animals.

Accordingly, the lack of response in most animal pathology studies did not lower the degree of certainty by much. Reviewers 1 and 3 had their degree of confidence increased somewhat by repeated but unreplicated results from one German laboratory (Mevissen et al., 1996b) and isolated results from two laboratories in the former Soviet Republics (Anisimov et al., 1996), (Beniashvili et al., 1991), which showed co-promotional effects on breast tumors. None of the reviewers were much influenced by the statistically significant increase in thyroid cancers in one of the bioassays (Boorman et al., 1999b), even though it had not appeared in control series of previous bioassays and was thus a very unlikely occurrence. This effect showed up in only one sex of rats and not in mice and thus did not pass conventional toxicological criteria for animal carcinogenicity.

BIOLOGICAL PLAUSIBILITY (MECHANISTIC STUDIES)

In setting their prior (initial degree of confidence), the reviewers already have discussed theoretical models based on general physics and biological knowledge, predicting that the threshold of possible influence above endogenous currents is higher than the environmental levels implicated by the epidemiological studies. They cannot, therefore, use this argument again with regard to new EMF-specific evidence. Various attempts were carried out as part of targeted EMF research to devise more refined models for the purpose of supporting or rejecting the hypothesis of an EMF risk. These are discussed in the section on mechanisms and therefore will not be re-evaluated each time the epidemiology of a specific endpoint is reviewed. The core evaluators thought that a lack of a definitive mechanistic explanation of how EMFs could induce biological change, or a chain of biological events leading to pathology, did not pull confidence down below its initial value. But neither did the chicken studies nor melatonin inhibition cell studies add much, if any, weight of evidence. They were, however, considered high priority for further study since they were relevant to the possibility of bioeffects at "low" levels of exposure.

ANALOGY

If a chemical with a particular structure causes cancer, one can argue by analogy that a similar chemical might have the same effect. The reviewers agree that analogy does not help much with the EMF issue. Many causal agents have no analogous situation to reason from, when first encountered, so the absence of an analogous agent does not pull their confidence down as much as the presence of a good analogous agent would pull them up. This situation does not vary from disease to disease.

TEMPORALITY

If one compared unemployment rates in the general population to those among prevalent cases of rheumatoid arthritis, one would see a higher unemployment among the arthritics. One would not conclude that unemployment causes arthritis because the above-mentioned study design has not ensured that the reviewers could rule out the possibility that the arthritis preceded the unemployment. The criterion of temporality simply requires that study designs rule out that kind of confusion. If they do not, then grave doubts would arise about the evidence. Confusions about temporality are not an issue in the EMF epidemiological study designs included in this evaluation. In an abundance of caution, the reviewers discuss and dismiss this issue in one of the miscarriage studies.

SPECIFICITY AND EVIDENCE FROM OTHER DISEASES

There is a tendency to believe specific associations between an agent and one disease or subtype of disease more than associations with more than one disease. This probably is because the likelihood of chance, bias, or confounding producing a false association with one specific disease or one subtype of, for example, cancer, is smaller than the likelihood of false associations with cancer type 1, 2, 3, or 4. But even with genotoxic carcinogens, more than one cancer may result from exposure. If an agent causes disease by perturbing the immune or endocrine system, the effects could be non-specific. The AIDS virus is associated with Kaposi's sarcoma in some cities and with lymphoma in others, apparently depending on the varying presence of other risk factors. EMFs are physical agents that reach all parts of the body and are not thought to work through traditional genotoxic mechanisms, if, indeed, they have a pathological effect. EMF associations have NOT been characterized by great specificity as to disease type or subtype. One's confidence in causality for disease X might be increased by one's confidence in causality for disease Y, particularly if they share common mechanisms or other features.

The core team members either gave no weight to lack of specificity or found that it increased the credibility (see the core team members' individual conclusions after each endpoint's evaluation).

COHERENCE/VISIBILITY

Sometimes the existence of one association logically suggests that another association also should hold true. When that happens, it is said that the evidence is coherent. For example, if maximum magnetic fields were associated with disease X, and electric blankets expose users to high maximum fields, then one would expect electric blankets to be associated with disease X. If sub-groups of the population are known to be more vulnerable to environmental insults, and EMFs are more strongly associated with disease X in the vulnerable group than in the non-vulnerable group, that, too, is an example of internal coherence.

While the discussion of the internal coherence of studies varied from endpoint to endpoint, the discussion of what is called "visibility" was valid for all diseases tracked by disease registries or reliably traceable through hospitalization records or death certificates.

When electrification came, initially to cities and then rural areas of the United States in the first half of the 20th century, each area went from zero to low average exposures and then to higher average exposures as electricity progressed from mere lighting to heating, cooking, and other uses. The reviewers would argue that personal exposure eventually may have fallen to somewhat lower exposures as affluence brought larger lot sizes, more underground lines, and less knob and tube wiring. But some have argued that the incidence of disease should have increased dramatically and linearly with increased production of electricity even though electricity use, as measured at the electric meter in a home or by kilowatts sold, is not necessarily associated with personal exposure to magnetic fields.

Some argue that, since we all are exposed to magnetic fields higher than those that preceded the introduction of electricity, there should be a change in disease rates over time and from places with more or less consumption of electricity. This assumes that even low levels of exposure cause substantial increases in risk. For the most part, the epidemiological associations have been with the top 5% or 10% of the exposed population. In Chapter 2 the reviewers provided calculations for the impact of various RRs conveyed by 95th percentile exposures. With relative risks below 3.00 this can be shown to produce less than a 15% fluctuation in the overall rate of disease. This size of an effect would be hard to disentangle from changes in other causes of the diseases in question. The reviewers discuss this in more detail in the chapters on childhood leukemia and spontaneous abortion, where there are associations between residential EMFs and disease. For spontaneous abortions and perhaps other diseases which are not routinely recorded and which usually are dealt with on an outpatient basis, larger impacts might have gone unnoticed. For the other diseases the reviewers take the generic position that the modest associations described might exist without being noticed as geographical or temporal fluctuations. They discuss the findings of Milham et al. (2002) with regard to electrification and childhood leukemia mortality in the chapter on that disease.

7.4 QUESTIONS RELEVANT TO POLICY

DOSE-RESPONSE QUESTIONS

Except for childhood leukemia and spontaneous abortion, there is not a sufficient evidentiary base or data to even speculate on the issues of thresholds, plateaus, special metrics, windows, and biological windows of vulnerability. The discussions of these topics are restricted primarily to the evidence from these two diseases.

RACIAL AND CLASS DIFFERENCES IN EXPOSURE AND VULNERABILITY

Policy perspectives that pay attention to environmental justice require evidence on special vulnerabilities or exposures. The reviewers discuss this in the chapter on exposure. With the exception of the two recent miscarriage studies sponsored by DHS, which found no racial or social class special vulnerability to EMFs, none of the papers they read presented data on potential differential impacts of EMFs on different races, ethnicities, or social class. This is noted in the summary tables.

HOW DOES THE OBSERVED RELATIVE RISK COMPARE TO THAT WHICH WOULD GENERATE A 1/100,000 OR 1/1000 LIFETIME ADDED RISK

Some regulatory frameworks consider as negligible (*de minimis*) those risks which would accumulate less than 1/100,000 added lifetime risk from 70 years of residential exposure or 1/1,000 during 40 years of occupational exposure. As an approximation, the reviewers took the crude mortality or incidence of the disease in question and applied the relative risk to obtain the annual theoretical incidence or mortality among "exposed" persons. They subtracted this number from 1.0 to obtain the probability of escaping that disease in one year. For 70 years of residential exposure, they raised that number to the 70th power to obtain the probability of escaping a particular disease in a lifetime. They then subtracted that from 1 to obtain the probability of contracting or dying from the disease in a 70-year lifetime. This was compared to the baseline lifetime probability of contracting or dying from that disease. A similar calculation was made for childhood cancer, but using 20 years, and for occupational cancers, using 40 years.

Epidemiological studies rarely have the resolution power to detect RRs less than 1.2 reliably. As a general rule, if the baseline incidence was equal to or greater than 1 per 100,000 per year, the reviewers determined that a RR of 1.2 or larger conveyed more than a 1/100,000 theoretical lifetime risk from 20 or 70 years of exposure. A baseline rate of 11/100,000 per year or greater was required if a 1.2 fold risk were to accumulate a 1/1,000 theoretical lifetime risk during 40 years of occupational life. This meant that all the agents would be of environmental regulatory concern if detectable by epidemiology. With a few exceptions (ALS, male breast cancer, adult brain cancer), they would be of regulatory concern in the workplace as well.

SIZE OF EMF RELATIVE RISKS AND ATTRIBUTABLE FRACTIONS COMPARED TO OTHER RISK FACTORS

Epidemiologists sometimes evaluate the "importance" of a factor by comparing the relative risk conveyed by the highest exposures and the proportion of the baseline rate due to this factor (the attributable fraction or PAR%) to those of other known factors. By these standards, cigarette smoking is large and exposure to other people who smoke is small when one considers lung cancer. The PAR% describes the expected percentage fall in the overall rate of the disease if the "exposure" were removed. It is a measure of effectiveness. But, at least in the utilitarian policy framework, it is cost effectiveness, not effectiveness, that guides priority setting. For example, highway speed accounts for most vehicular injury fatalities, but the economic and political cost of enforcing a 25 mile-per-hour speed limit (or even a 55 mile-per-hour speed limit) on the freeway makes that strategy less cost effective than enforcing the use of seatbelts. Nonetheless, since the PAR% is a criterion often used, the reviewers address it in the structured questions.

7.5 WHY CANCER CLUSTER LITERATURE IS NOT REVIEWED

Although public and media attention to the EMF issue has been stimulated in great part by reports of cancer clusters near power lines or transformer stations, as well as radio frequency and radar transmitters, the DHS reviewers have not (nor have the NIEHS, NAS, and WHO) included a review of these reports. The reason is that this stream of evidence for EMFs carries little weight. Even if EMFs increase the risk of certain cancers, the proportion of neighborhoods displaying a cancer cluster above what was expected would be low (the test is not "sensitive"). For example, in Sweden, Feychting and Ahlbom (Feychting & Ahlbom, 1993) identified all childhood cancers that had occurred over many decades within 300 meters of the thousands of miles of transmission lines. By accumulating all this information they identified an excess number of childhood leukemia cases within 50 meters of the line. The excess was a handful of cases spread along the many miles of transmission line which ran through inhabited areas. There were not enough cases in those many decades to form a cluster that any neighborhood group would have noticed.

But cluster evidence generates false positives, that is, it is not "specific." This can be predicted by the laws of probability. Since the California Cancer Registry routinely tracks 50 kinds of cancer, the chance that any one suburban city block will escape a statistically significant ($p = .01$) elevation of all these 50 cancers is 0.99 to the 50^{th} power or 60%. That means there is a 40% probability that at least one of those 50 cancers will be found in excess. Inasmuch as the approximately 10 million California households are grouped in a few 100,000 blocks and about 2% of those blocks are near enough to transmission lines to influence the magnetic field levels (Lee et al., 2000), 40% of a few thousand blocks near transmission lines would be found to have at least one of those 50 kinds of cancer, by chance alone (Neutra, 1990).

If one wanted to examine clusters as a legitimate test of the EMF hypothesis, one would examine the 1,000 or so city blocks near transmission lines and compare the number of cancer clusters on them to the number on a 1,000 blocks of similar socioeconomic status but away from transmission lines. The vast majority of the clusters would be from the 40% of blocks with chance clusters and a few extra clusters might be detected if the nearby lines were a causative agent. The strategy of Feychting (1993) is a better strategy because it pays attention to all the cancers, not just the ones which occur in clusters. It is for this reason that the reviewers restrict their examination to well-designed epidemiological studies.

7.6 HEURISTIC FOR UPDATING THE DEGREE OF CONFIDENCE IN CAUSALITY

The ideal way to develop a posterior degree of confidence would be to develop a full probabilistic model or Bayesian Net, but the reviewers' stakeholders made clear at the outset that they should not rely on a method that would not be accessible for criticism to most readers.

Accordingly, the reviewers have structured their narrative to reflect the considerations that would go into a Bayesian net and elicited their posterior degrees of confidence directly after systematically considering the narrative. The reviewers used numbers, as well as the agreed-upon everyday language phrases, to characterize their professional judgments. They also applied the IARC criteria to derive a categorization of the evidence according to traditional guidelines.